Online og frit tilgængelige genomiske analyser

Bioinformatics for dummies

Frank M. Aarestrup

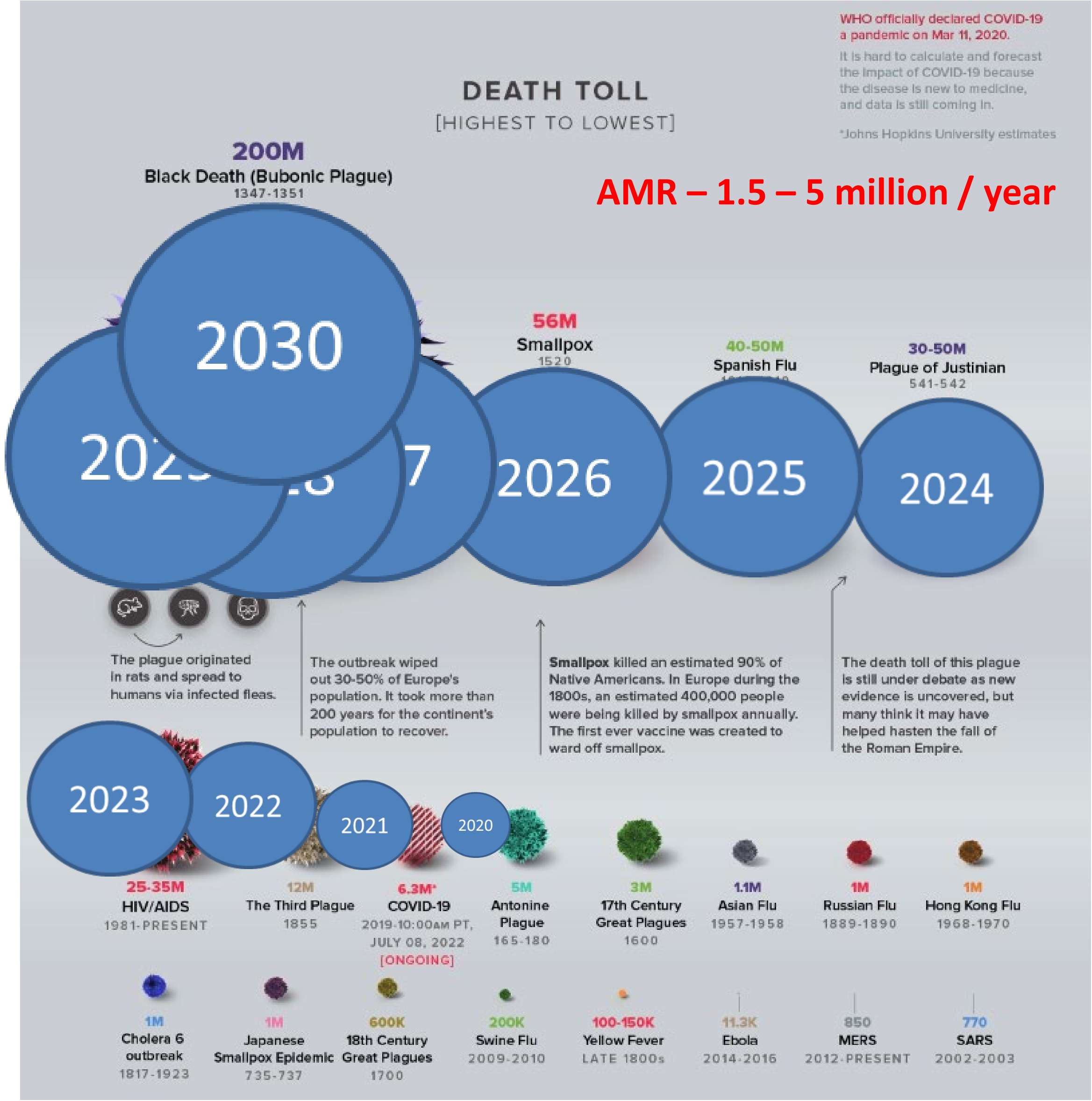
fmaa@food.dtu.dk, www.genomicepidemiology, www.globalsurveillance.eu

Infektionssygdomme er en af vore største sundhedsmæssige problemer

- Mange forskellige patogener
  - Oversete epidemier (*E. coli*, *Staphylococcus*)
  - udbrud (Salmonella, cholera, listeria)
  - Pandemier (influenza, corona, cholera) – **COVID-19 en af mange**

- Enorme konsekvenser
  - Morbidity, mortality, economic

- Antibiotika er ofte essentielle



WHO officially declared COVID-19 a pandemic on Mar 11, 2020.

It is hard to calculate and forecast the impact of COVID-19 because the disease is new to medicine, and data is still coming in.

*Johns Hopkins University estimates

AMR – 1.5 – 5 million / year

The plague originated in rats and spread to humans via infected fleas.

The outbreak wiped out 30-50% of Europe's population. It took more than 200 years for the continent's population to recover.

Smallpox killed an estimated 90% of Native Americans. In Europe during the 1800s, an estimated 400,000 people were being killed by smallpox annually. The first ever vaccine was created to ward off smallpox.

The death toll of this plague is still under debate as new evidence is uncovered, but many think it may have helped hasten the fall of the Roman Empire.

Hvad verden behøver

- Real-tid data over forekomst og spredning af alle patogener og deres resistens
  - Uafhængig af geografi, reservoir, og patogen

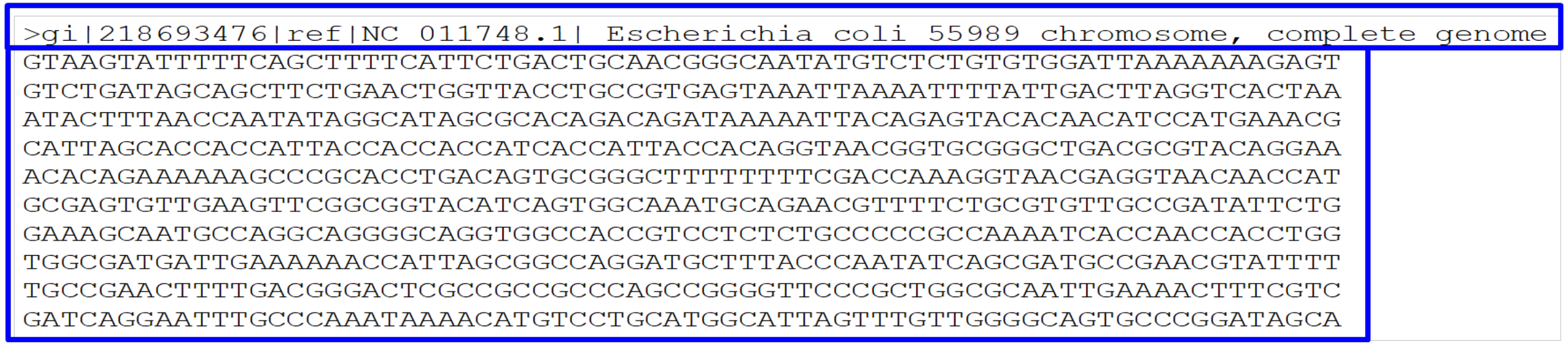
  - Mulighed for hurtige analyser og deling af data:
    - Myndigheder
    - Udvikling af test og vacciner
    - Kliniske beslutninger

Next Generation Sequencing (NGS)
Fordele

- DNA/RNA er fælles for alle patogener og metoder til analyse an derfor være fælles

- Det giver et fælles sprog

Sequence data:



```
>gi|218693476|ref|NC 011748.1| Escherichia coli 55989 chromosome, complete genome
GTAAGTATTTTTCAGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGT
GTCTGATAGCAGCTTCTGAACTGGTTACCTGCCGTGAGTAAATTAAAATTTTATTGACTTAGGTCACTAA
ATACTTTAACCAATATAGGCATAGCGCACAGACAGATAAAAATTACAGAGTACACAACATCCATGAAACG
CATTAGCACCACCATTACCACCACCATCACCATTACCACAGGTAACGGTGCGGGCTGACGCGTACAGGAA
ACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTCGACCAAAGGTAACGAGGTAACAACCAT
GCGAGTGTTGAAGTTCGGCGGTACATCAGTGGCAAATGCAGAACGTTTTCTGCGTGTTGCCGATATTCTG
GAAAGCAATGCCAGGCAGGGGCAGGTGGCCACCGTCCTCTCTGCCCCCGCCAAAATCACCAACCACCTGG
TGGCGATGATTGAAAAAACCATTAGCGGCCAGGATGCTTTACCCAATATCAGCGATGCCGAACGTATTTT
TGCCGAACTTTTGACGGGACTCGCCGCCGCCCAGCCGGGGTTCCCGCTGGCGCAATTGAAAACTTTCGTC
GATCAGGAATTTGCCCAAATAAAACATGTCCTGCATGGCATTAGTTTGTTGGGGCAGTGCCCGGATAGCA
```

2009 vi kan alle dele

Behov

- Hurtigere deling af data

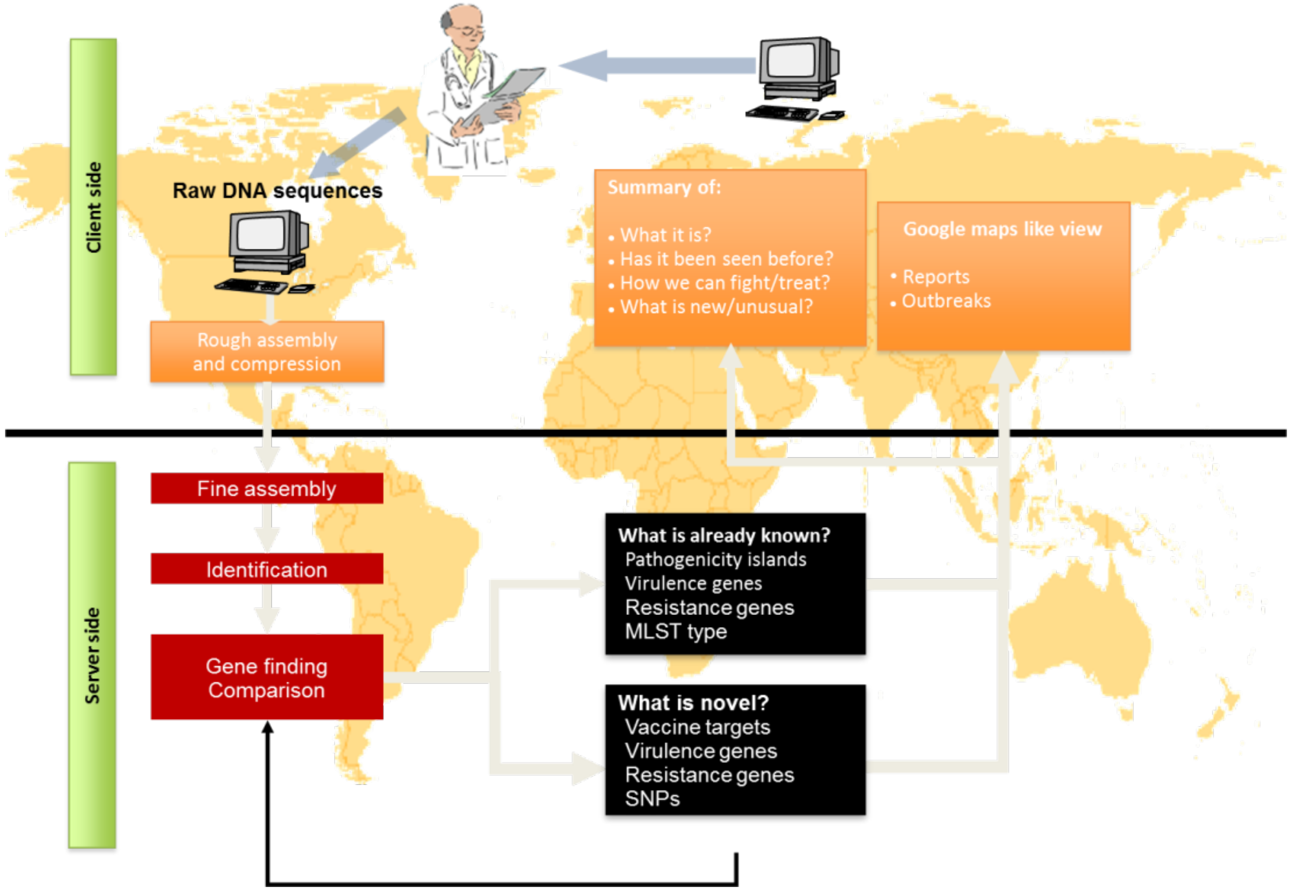
Løsning

- NGS er et fælles sprog der nemt og hurtigt kan deles

Forskning

- Bioinformatics for dummies

Og vi levede lykkeligt til vore dages ende



Client side

Raw DNA sequences

Rough assembly and compression

Summary of:
- What it is?
- Has it been seen before?
- How we can fight/treat?
- What is new/unusual?

Google maps like view
- Reports
- Outbreaks

Server side

Fine assembly

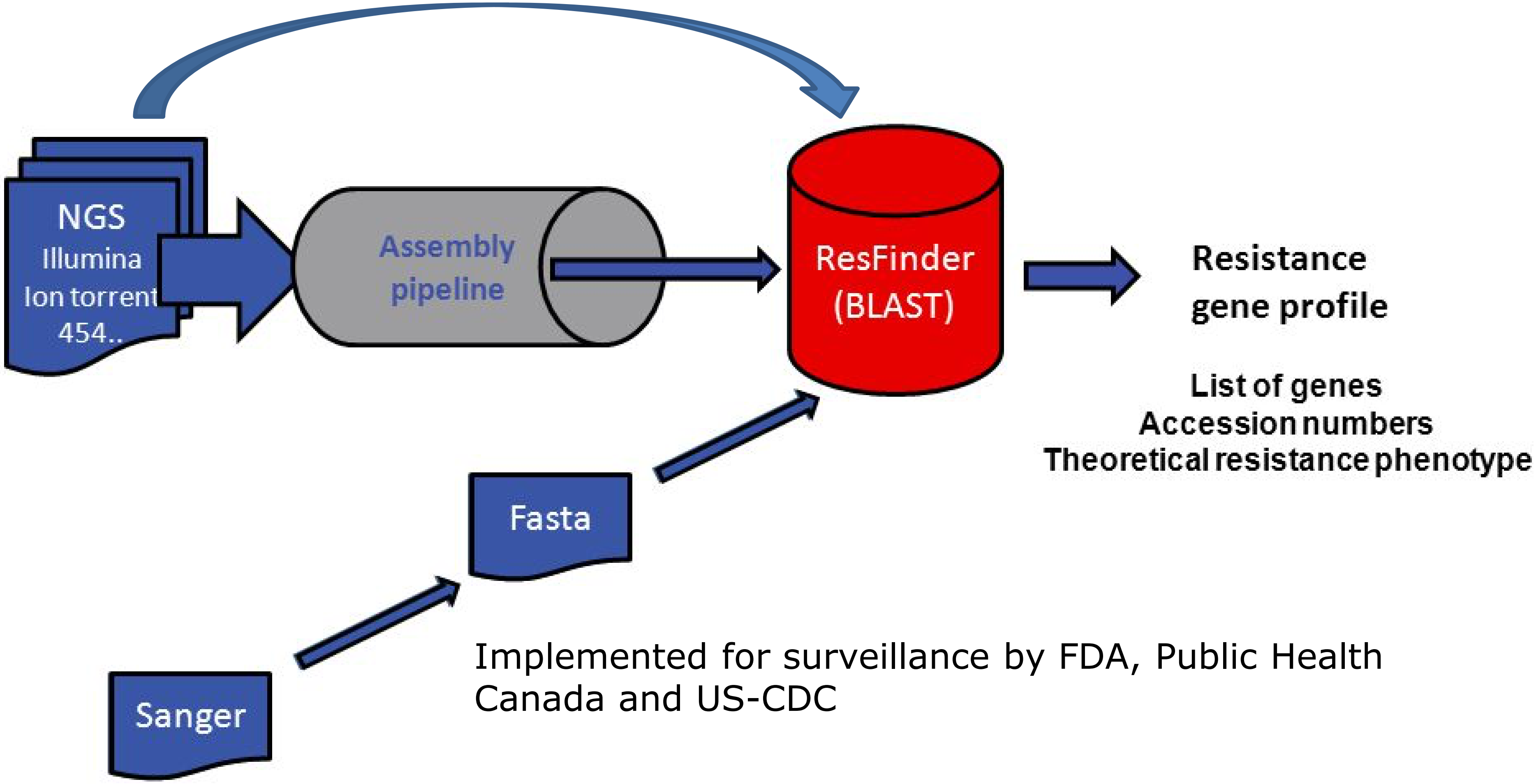
Identification

Gene finding
Comparison

What is already known?
Pathogenicity islands
Virulence genes
Resistance genes
MLST type

What is novel?
Vaccine targets
Virulence genes
Resistance genes
SNPs

ResFinder

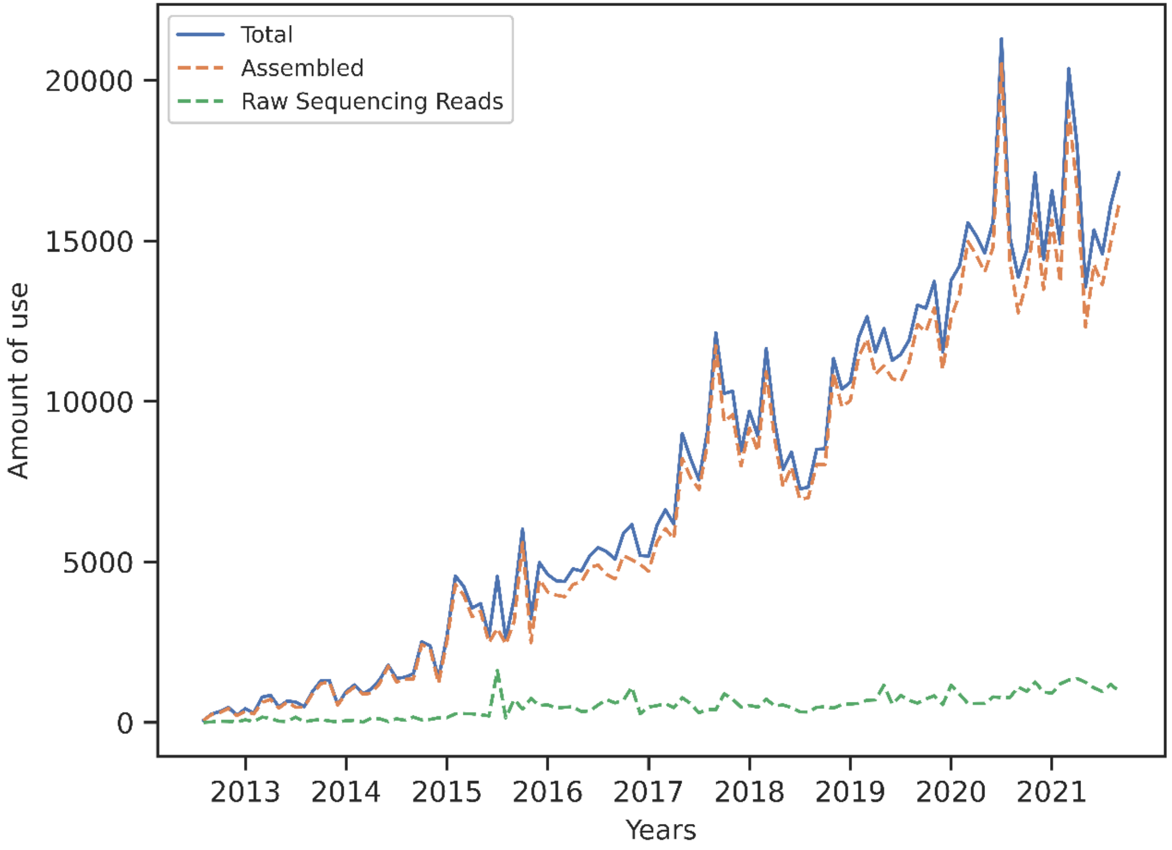


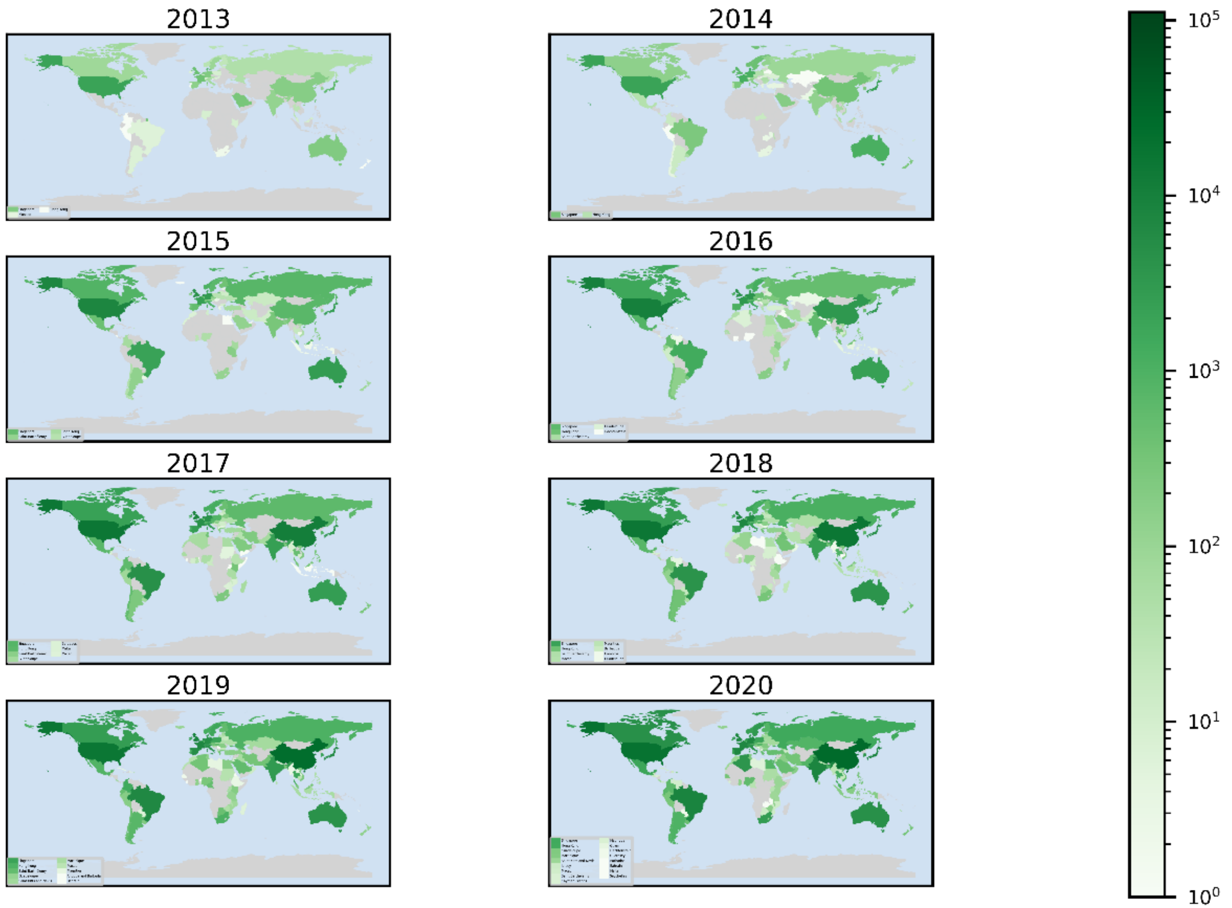
Resistance gene profile

List of genes
Accession numbers
Theoretical resistance phenotype

Implemented for surveillance by FDA, Public Health Canada and US-CDC

ResFinder 4.0 provides predicted phenotypes





Current use is around 30,000 analyses per month and LMICs have started to become visible

2021 – Deler vi så globalt?

Behov

- Hurtigere deling af data

Løsning

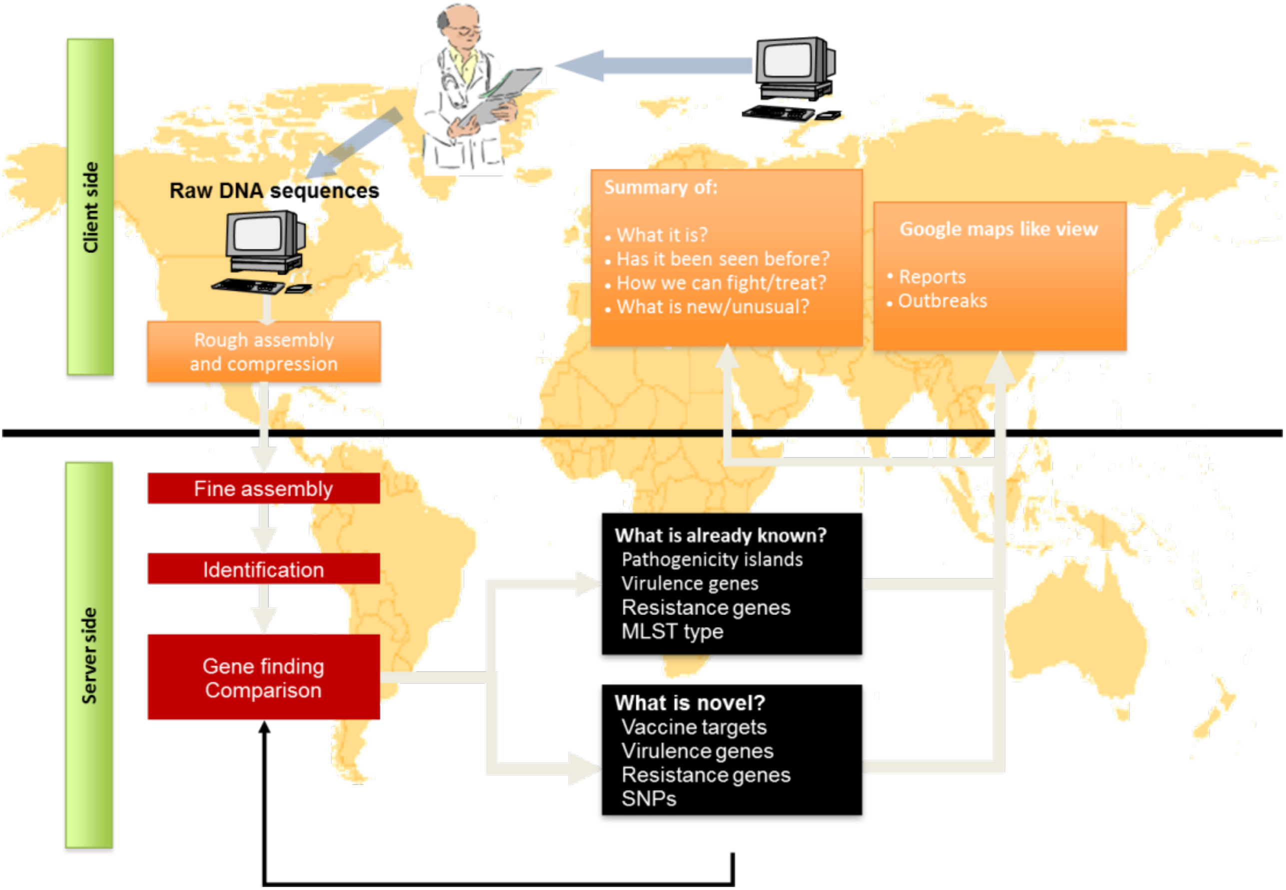
- NGS er et fælles sprog der nemt og hurtigt kan deles

Forskning

- Bioinformatics for dummies

Og vi levede lykkeligt til vore dages ende

Stadig primært et forskningsværktøj



Client side

Raw DNA sequences

Rough assembly and compression

Summary of:

- What it is?
- Has it been seen before?
- How we can fight/treat?
- What is new/unusual?

Google maps like view

- Reports
- Outbreaks

Server side

Fine assembly

Identification

Gene finding Comparison

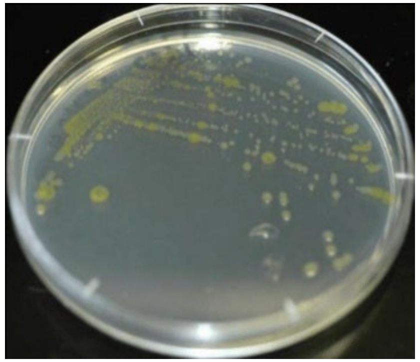
What is already known?
Pathogenicity islands
Virulence genes
Resistance genes
MLST type

What is novel?
Vaccine targets
Virulence genes
Resistance genes
SNPs

Microbial Outbreak Surveillance System (MOSS)

Pathogen sampling from patient and cultivation

Over night

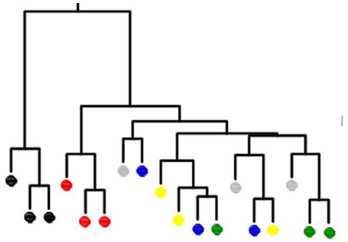


Purification
And
Library prep.

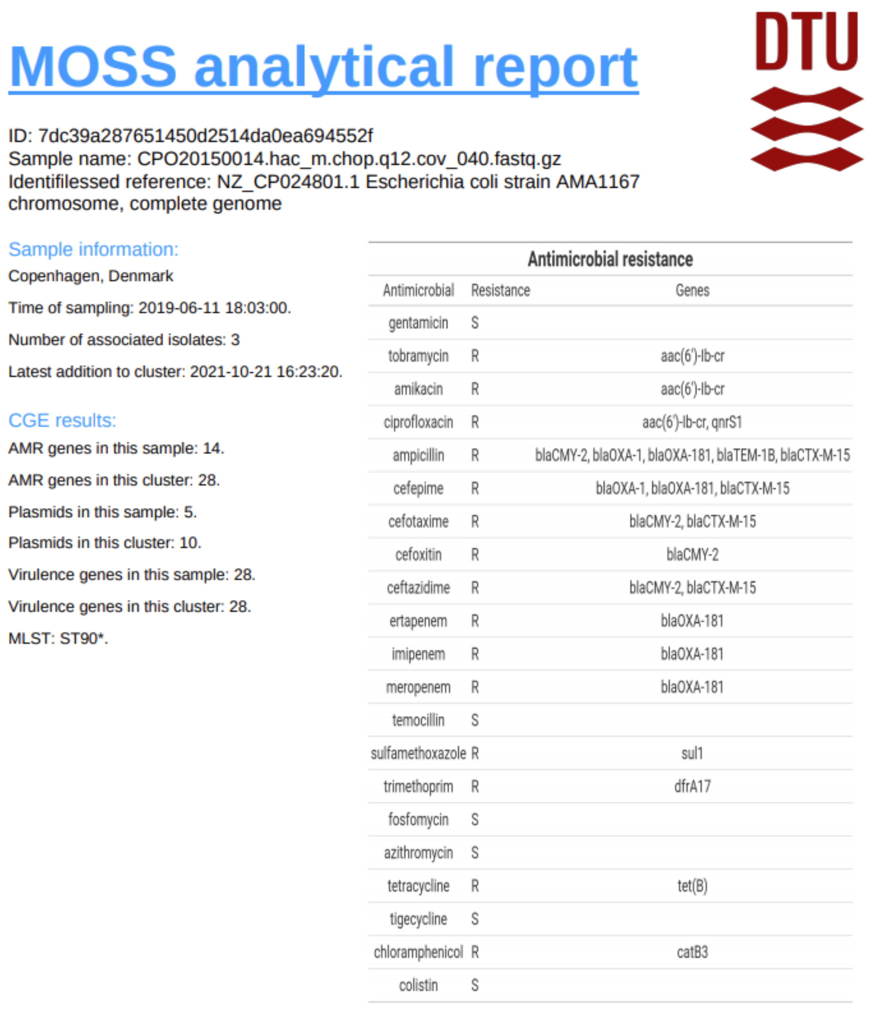
2 - 8 hours

Sequencing

0.5 - 3 hours



Analysis
~5min



MOSS analytical report

ID: 7dc39a287651450d2514da0ea694552f
Sample name: CPO20150014.hac_m.chop.q12.cov_040.fastq.gz
Identifilessed reference: NZ_CP024801.1 Escherichia coli strain AMA1167 chromosome, complete genome

Sample information:

Copenhagen, Denmark

Time of sampling: 2019-06-11 18:03:00.

Number of associated isolates: 3

Latest addition to cluster: 2021-10-21 16:23:20.

CGE results:

AMR genes in this sample: 14.

AMR genes in this cluster: 28.

Plasmids in this sample: 5.

Plasmids in this cluster: 10.

Virulence genes in this sample: 28.

Virulence genes in this cluster: 28.

MLST: ST90*.

Antimicrobial resistance

Antimicrobial	Resistance	Genes
gentamicin	S	
tobramycin	R	aac(6')-Ib-cr
amikacin	R	aac(6')-Ib-cr
ciprofloxacin	R	aac(6')-Ib-cr, qnrS1
ampicillin	R	blaCMY-2, blaOXA-1, blaOXA-181, blaTEM-1B, blaCTX-M-15
cefepime	R	blaOXA-1, blaOXA-181, blaCTX-M-15
cefotaxime	R	blaCMY-2, blaCTX-M-15
cefoxitin	R	blaCMY-2
ceftazidime	R	blaCMY-2, blaCTX-M-15
ertapenem	R	blaOXA-181
imipenem	R	blaOXA-181
meropenem	R	blaOXA-181
temocillin	S	
sulfamethoxazole	R	sul1
trimethoprim	R	dfrA17
fosfomycin	S	
azithromycin	S	
tetracycline	R	tet(B)
tigecycline	S	
chloramphenicol	R	catB3
colistin	S	

In-depth analysis of clusters available in front-end client, but summaries are also complied into a PDF report for each sample analyzed.

Felt-studium i Tanzania

- Dreng med diarre -> Campylobacter -> Ændret behandling -> respons og hjemsendelse

NGS, open science, data-deling og global overvågning

- Store potentialer i NGS

- Online service bruges intensivt, men stadig mest til forskning

- Behov for at integrere den primære diagnostik mere

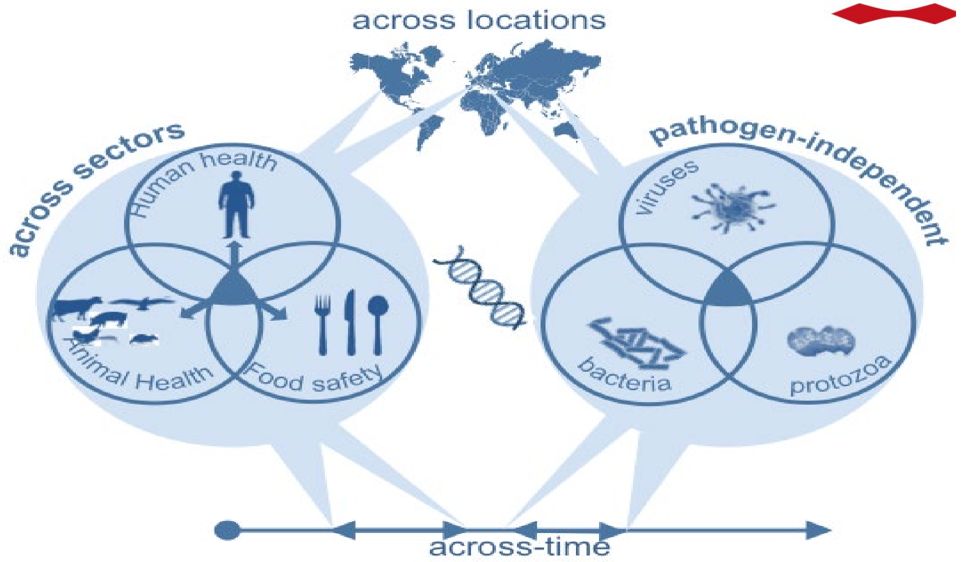
  - Videreudvikling af bioinformatikken
  - Nemmere deling af data
  - Face-book for genomics

Our vision: one system serves all

DTU

Guiding principles:

- Cross sector, cross domain, open source (not commercial)
- Interaction with the rest of the world (all inclusive)
- Data for action (actionable outputs)
- Central repository (tools to the data) (ENA, DDJ,NCBI)



The
Danish Council for
Strategic Research

DANIDA
Ministry of Foreign
Affairs of Denmark

efsa
European Food Safety Authority

VILLUM FONDEN

WHO Collaborating Centre
for Antimicrobial Resistance in
Foodborne Pathogens and Genomics
www.antimicrobialresistance.dk

HORIZON 2020

European
Commission

novo nordisk fonden